DOCUMENT RESUME

ED 435 404 IR 057 554

AUTHOR Mardis, Kathleen P.

TITLE Here Today, Gone Tomorrow: The Stability of Cited Electronic References.

PUB DATE 1999-05-00
NOTE 35p.; Master's Research Paper, Kent State University.
PUB TYPE Dissertations/Theses (040)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Citations (References); Electronic Text; *Error Patterns; Information Retrieval; Information Science; *Internet; Listservs; Scholarly Journals; Tables (Data)
IDENTIFIERS *Electronic Resources; File Transfer Protocol; Gopher; URLs; Web Sites

ABSTRACT

Numerous studies of the accuracy of citations have appeared for a variety of disciplines including library and information science. To date, these studies have focused on the print format of the professional literatures. It is only recently that studies have been performed on the professional literature available electronically, specifically Internet resources. The research conducted thus far has primarily addressed the growing body of electronic journals and related issues of accessibility and retrieval. Preservation issues have also been discussed. What have not been examined are citations to documents available on the Internet that are increasingly being referenced in the research literature. How stable are electronic resources over time? What percentage of errors occurs in electronic references? What types of references (i.e., HTTP, FTP, Gopher, Listservs) are most stable over time? What types of errors exist in citations? Do the number of errors for a given type increase over time? In this study, all citations to electronic resources for 10 library and information science journals for 1994-98 were examined. The total number of retrieved citations was calculated by year and type. Errors were sorted by type and classified as major or minor. It was found that the average rate of stability of all electronic resources for the time period examined was 55.1%. (Contains 29 references and 12 tables.) (Author/MES)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 435 404

HERE TODAY, GONE TOMORROW: THE STABILITY OF CITED ELECTRONIC REFERENCES

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY

D.P. Wallace

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

A Master's Research Paper submitted to the
Kent State University School of Library
and Information Science
in partial fulfillment of the requirements
for the degree Master of Library Science

by

Kathleen P. Mardis

May, 1999

IR057554

Numerous studies of the accuracy of citations have appeared for a variety of disciplines including library and information science. To date, these studies have focused on the print format of the professional literatures. However, it is only recently that studies have been performed on the professional literature available electronically, specifically Internet resources. The research conducted thus far has primarily addressed the growing body of electronic journals and related issues of accessibility and retrieval. Preservation issues have also been discussed. What have not been examined are the citations to documents available on the Internet that are increasingly being referenced in the research literature. How stable are electronic resources over time? What percentage of errors occurs in cited electronic references? What types of references are most stable over time? What types of errors exist in the citations? And do the number of errors for a given type increase over time? In this study, all citations to electronic resources for ten library and information science journals for the years 1994-1998 were examined. The total number of retrieved citations was calculated by year and type. Errors were sorted by type and classified as major or minor errors. It was found that, on average, the rate of stability of all electronic resources, for the time period examined, was 55.1%.

Master's Research Paper by

Kathleen P. Mardis

B.A., The Ohio State University, 1981

B.S. Ed., The Ohio State University, 1990

M.L.S., Kent State University, 1999

Approved by

Adviser Lechera N. Connell Date April 26, 1999

TABLE OF CONTENTS

LIST OF TABLES.....iv

INTRODUCTION.....1

LITERATURE REVIEW.....3

DEFINITIONS.....9

RESEARCH QUESTIONS.....11

METHODOLOGY.....12

RESULTS AND DISCUSSION.....14

CONCLUSIONS.....24

BIBLIOGRAPHY.....26

LIST OF TABLES

Table 1: Minor error types....14

Table 2: Major error types....14

Table: 3 Total citations retrieved by year....16

Table: 4 Total citations retrieved by reference category....16

Table: 5 Retrieval rate by category over time....17

Table 6: Total errors by resource category....18

Table 7: Total errors over time....19

Table 8: Minor errors by citation category and type of error....20

Table 9: Minor errors by type over time....21

Table 10: Major errors by citation category and type of error....22

Table 11: Major errors by error type over time....23

Table 12: Comparison of two classes of major errors over time....24

ACKNOWLEDGEMENTS

I would like to thank my research adviser, Dr. Tschera H. Connell for all of her assistance, support, good humor, and the idea for this paper without which I would have been lost. I would also like to thank Dr. Mary T. Kim and Dr. Mary Stansbury. ...just cuz.

Introduction

The standardization of citations in the professional literature and the preservation of that literature have long aided scholars in the research process. The procedure for locating and accessing cited references has relied on this standardization. And, the process of preserving and archiving that literature has ensured that the information is available when needed. The advent of the Internet and web publishing has brought some profound changes to the way we locate and access information. Yet, the use of a single standard is still needed for identifying, labeling, and preserving resources that are located on the Internet for future retrieval. The International Organization for Standardization has released ISO 690-2 outlining requirements for bibliographic references of electronic resources. (ISO 1998) However, this standard is not used consistently in the literature in which citations to electronic resources are making an increasing appearance. References to papers, articles, electronic journals, as well as discussion lists, bulletin board postings, and email messages are being cited along with the traditional references. The use of multiple citation styles creates potential problems in retrieving these cited items.

Likewise archival procedures are needed to ensure the continued existence of these newest resources as it becomes increasingly apparent that web pages which were present a year ago are now often unavailable. Librarians and researchers have grown frustrated at not being able to locate pages that they had carefully bookmarked for future reference. "Although it is human nature to be a little slow on the uptake when it comes to preserving our contributions to

the world record ... We lament our ancestors' lack of insight for not preserving things they should have recognized as important, and yet we suffer this same affliction. (Casey 1998, p. 304) The professional literature of a given discipline, and the citations to that literature are certainly worthy of preserving. However, the creation of cyberarchives has not yet become a practice of research institutions.

Despite their similarities, there are significant differences that merit attention between print and electronic resources. Unlike their print counterparts, electronic references lack permanency. The very nature of the medium implies that these resources may be located and/or relocated anywhere on the Internet, which means anywhere in the world. Printed resources are not subject to change through relocation. An article that appears in a given issue of journal "A" will not one day disappear and then reappear in journal "B". In addition, electronic documents are subject to repeated updating in which content is changed but not necessarily identified as revised. This practice, albeit rare in the print literature, produces a new document that is identified as a revision. As a result, a researcher may cite a reference to an electronic document which later may not exist in the same version he used. The question of whether a given document exists, as cited, then becomes enigmatic.

Finally, the citations to print literature may contain abundant errors or inaccuracies. However, it is usually possible to locate and access the cited print references. Electronic references are located and accessed using a Uniform Resource Locator (URL), a Usenet news group address, or Listserv address. All

of these identifiers are subject to change due to new host server configuration, changes in file structure, or the purging of archive news postings. Finding these resources then becomes tantamount to finding a small needle in a very big haystack; the only clues given the searcher are "404 file not found " or "this message was undeliverable."

The stability of web sites, more often than not, depends on who owns and controls the host server on which they are located. While it is can be asserted that not everything currently available on the Web needs to be preserved, it would seem to be essential to preserve the professional literature being cited in whatever media it exists. However, the question then becomes, who is responsible for preserving and providing access to the literature found on these web sites.

Literature review

Citation referencing represents the "scientific bricklaying" of the research process, linking an earlier work to a later one. (Mitra 1970) The importance of this "bricklaying" has been outlined numerous times. Garfield offers a list of fifteen possible reasons for citations (Garfield 1979). His list includes "paying homage to pioneers and giving credit for related work (homage to peers)." Baird and Oppenheim (1994) also discuss numerous reasons people cite. Their reasons include: to criticize an earlier paper or argument, to present background to the current topic, to offer corroboration for one's argument or ideas, to identify the original source of one's current ideas, and to cite "a major figure because it

makes your research look more respectable" (Baird and Oppenheim 1994, p. 6). Mitra (1970, p. 117) stresses the importance of bibliographical referencing as providing " the necessary 'currency' for recognition of a particular scientist among his peers as well as for establishing his property rights and priority claims with respect to the scientific contribution he makes." These reasons, as well as others not listed, provide ample justification for the need to maintain stable and accurate citations in the literature.

The current practice of explicit bibliographical referencing through citations began around 1850. (Price 1963) The format of these citations evolved over the subsequent decades, from earlier footnote structure to the formats detailed in the various style manuals of the 1950's. Citation style guides for electronic resources are beginning to appear which offer standard formats for the different types of electronic resources found on the Internet. Some institutions such as Griffith University in Australia have set up depositories for documents dealing with the emerging standards for electronic resources and the scholarly citations of these resources. (Greenhill 1998) Currently, electronic citation guides are available online for APA (Publication Manual of the American Psychological Association) and MLA (MLA Handbook for Writers of research Papers) styles. (Guffey 1997, Whitley 1997, Li and Crane 1996)

Although the various style guides do provide some level of "standardization" for bibliographical referencing, they are insufficient to establish an authoritative standard. The International Organization for Standardization is such an authoritative body and has released its own specifications for

bibliographic references of electronic resources. (ISO 1998) The International Standard ISO 690-2 specifies the required data elements and their order of appearance in references for electronic documents and "establishes conventions for the transcription and presentation of information derived from the source electronic document." (ISO 1998) The ISO 690-2 standard outlines the format for electronic documents as well as electronic messages located online and provides example citations for these types of references.

The growing impact of electronic resources on scholarly literature has been examined and citation-based indicators for impact measurement have been proposed. (Zhang 1998, p. 246) One of the conclusions of this study pointed to the lack of "citing conventions" for electronic sources. Many of the journals included in this study do not have specific editorial guidelines for authors to use when citing electronic sources. (Zhang 1998, p. 249) In addition, this study was supported by the work of Harter and Kim which found that electronic journal articles are more likely to cite electronic resources than print articles. However, there was no significant difference observed in the number of electronic references by journal format. (Harter and Kim 1996a)

Harter, Kim and Ford have also studied the accessibility of electronic journals. (Harter and Kim 1996b; Ford and Harter 1998) They looked at such issues as the accuracy of URLs listed in directories and catalogs and determined whether or not they led to viable sites; the problems users encounter in accessing e-journals; and the existence and/or creation of electronic archives. These studies can be used to support the assertion that electronic resources are

growing in importance in the professional literature and continued access to these sources is needed.

While standardization of bibliographical references for electronic resources is necessary, unique identifiers for these resources are also required. The responsibility for the development of Internet identifiers has resided primarily with the Internet Engineering Task Force (IETF) which has been charged with the task of defining the needed infrastructure for the Internet. (The IETF was formerly under the direction of the Internet Architecture Board which is responsible for defining the overall architecture of the Internet.) As early as 1992, the IETF set up a working group to develop Uniform Resource Identifiers (URIs) as a standardization for the various URL formats that were appearing on the Internet. (Berners-Lee 1994) The model developed by this group called for the use of Uniform Resource Names (URNs) "which would be a persistent identifier for a resource, independent of information such as protocol, host, port, etc." (Daigle 1998, p. 329)

URNs do not refer to a specific location on the Internet but to a unique resource. A URN may identify "intellectual content, or whatever a name assignment authority determines is a distinctly namable entity." (Sollins and Masinter 1994) Another requirement for a URN is its persistence. The assignment of a URN is permanent. The URN "may well be used as a reference to a resource well beyond the lifetime of the resource it identifies." (Sollins and Masinter 1994) It is logical to assume that these URNs might be used in bibliographical citations as part of a "standardization". To date, that hasn't

happened. The use of URNs has yet to become a universal reality.

To assist the continued development of the uniform resource naming system, OCLC (Online Computer Library Center), in 1996, endeavored to create a resolution service through the use of Persistent URLs (PURLs). The principle behind PURLs is that a resolution service creates and maintains an intermediate association between an Internet resource and its URL, also known as an HTTP redirect. (Shafer, Weibel, Jul and Fausey 1996; Weibel, Jul and Shafer 1996) PURLs are needed because Internet resources move, change their names, or their mode of access. Once a URL changes, all links to that URL become invalid.

PURLs are used to assign a "name" for an electronic resource that is persistent regardless of the location, or change in location of that resource. If a web page is moved to a new location, a PURL will continue to reference that web page. As a result, bibliographic references remain viable over time. (Library of Congress 1997) The OCLC team points out that "persistence is a function of organizations, not technology." (Shafer, Weibel, Jul and Fausey 1996) Those resources that require long-term access are excellent candidates for PURLs. Such resources might include electronic journals, articles, and conference papers. However, use of a PURL resolver service is not an automatic process. A "maintainer" is employed to update the URL associations. These maintainers are members of a group of registered users of the PURL system. The PURL system is useful for maintaining accessibility of web pages. However, it is not a system for archival preservation of those pages.

An institutional approach to URL stability has been developed by the University of Waterloo (Canada) in their Scholarly Societies Project. The Scholarly Societies Project has developed a URL-Stability Index for the 200 URLs that are used for the Project website. This group has created a "canonical domain-name format" to resolve the conflict of changing URLs. (Parrott 1998) A domain name is defined as "a 'permanent' address assigned to an institution or group.'" (Parrott 1998)

The stability index is defined by assigning a value to a URL based on the format of the URL. Canonical domain-name URLs are permanent URLs and are reserved for resources that should not change over time. These URLs are assigned a value of 1. URLs with no domain name are assigned a value of 0. These URLs are then tracked over time. The Scholarly Societies Project also evaluated subject disciplines in the university based on the URL-Stability index.

The need for standardization is well documented and tested in the literature. Numerous citation studies have been conducted examining the citation errors and inaccuracies of bibliographical references. In an often cited study of bibliographical citations, Sweetland suggests that citation errors can be attributed to

> a lack of standardization in citation formats, misunderstanding of foreign languages, general human inabilities to reproduce long strings of information correctly, and failure to examine the document cited, combined with a general lack of training in the norms of citation. (Sweetland 1989, p. 291)

Citation studies examining the errors in references in the library science literature have concluded that information professionals commit the same mistakes in citations as are found in other disciplines (Pandit 1993). Pandit examined the citation errors in five library science journals and found that most errors related to numerical data; errors occurred in volume numbers, issue numbers, page numbers and publication years.

Pope also studied the library science literature. (Pope 1992) She examined a random sample of citations from ten library science journals and her results were comparable in other professional literatures. Pope concluded that "format and content of citations not only varied according to the [publishing] guidelines of the journals, but also within a journal." (Pope 1992, p. 242)

These studies suggest that a standardized format of citations is essential. Yet the existence of a standard is no guarantee of the accuracy of a citation or the probability of locating and retrieving a referenced source.

Definitions

Electronic reference is used here to mean any resource located on the Internet as cited in a journal article. An article is defined as a research paper appearing in a journal. For the purpose of this study, electronic reference and electronic resource are synonyms. A citation is defined as "a reference to some previously published work that is relevant to the argument the author wants to make." (Baird and Oppenheim 1994, p. 3)

Errors or discrepancies are defined as "deviations from the source" (Doms

1989, p. 442) and include inaccurate or incomplete information in the citation. Doms distinguishes between major and minor errors in cited references. She defines minor errors as "omissions that did not prevent the location of the article" and major errors as "references that prevented immediate location of the article cited" (Doms 1989, p. 442). This study distinguishes minor versus major errors according to Doms's definitions.

Stability is defined here as the access permanence of cited electronic resources over time. Uniform Resource Locators (URLs) are "the 'addresses' of resources on the WWW (World Wide Web)[and provide a] 'compact representation' of the location and access method for a resource available on the Internet." (Daigle, Daniel and Preston 1998, p. 327-328) Uniform Resource Identifier (URI) is "the generic name for a class of identifying tags of labels'" and Uniform Resource Name (URN) is "an identifying tag for files independent of server location." (Hoemann 1995) Within the organization of the Internet, "URNs are used for identification, and URLs for locating or finding resources." (Sollins and Masinter 1994).

The categories of resources examined in this study include three kinds of URLs: Hyper Text Transfer Protocol (HTTP), File Transfer Protocol (FTP), and Gopher, and Listserv messages. Listservs are automated mail distribution systems for which posted messages are distributed to subscribers of a given list of topical interest. Cited references to postings of news groups and personal email messages do appear in the literature however, they were not included in this study. For the purposes of this study, it was assumed that the researcher

would be able to access the electronic resources cited from the Internet. Any resource with restricted access, e.g., a required password, was included and noted in the study. It was also assumed that all information needed to access the resource would be found in the citation.

Research questions

To date, no citation analysis of cited electronic references has been published in the literature. This study examined ten journals in the field of library and information science and attempted to answer the following research questions:

1. What is the most stable category of electronic resource?
2. What percentage of errors occurs in cited electronic references? How does that compare to citation errors for traditional print references in the field of library and information science?
3. Does the percentage of errors increase over time?
4. Which electronic resources, e.g., FTP sites, have the highest number of errors? Does the number of errors for each category of electronic resource increase over time?
5. What types of errors can be identified?

Methodology

In order to evaluate the stability of cited electronic resources, a citation analysis was conducted. Ten journals in the field of library and information science for the period 1994 to 1998 were selected for this study.

Studies of citation error have employed a variety of sampling methods, (Moed and Vriens 1989, Pope 1992, Pandit 1993). However, the recent appearance of cited electronic resources in the scholarly literature led this researcher to conclude that they currently represent only a small fraction of the total number of citations. To answer the research questions, to improve reliability, and to insure an adequate sample size, all citations to electronic resources found at the end of journal articles were examined. A total of 1157 citations were examined.

For this study, the ten journals were selected after reviewing the 1996 ISI's Journal Citation Reports (SSCI 1996) ranked list of journals. The journals chosen are among those ranked 25 and higher within the category for information science and library science. Additional criteria for selection were: the requirement that the journal published primarily peer-reviewed research articles as opposed to informative articles for the general public (e.g., *Online*) or review articles (e.g., *Annual Review of Information Science and Technology*). The final criterion was the availability of the original publication issues for the selected journals and their current subscriptions held by three academic institutions in the state (Kent State University, The Ohio State University and Wright State University). The journals selected for this study are: *Journal of the American*

Society for Information Science, Journal of Documentation, College & Research Libraries, Information Processing & Management, Scientometrics, Library & Information Science Research, Library Quarterly, Interlending Document Supply, Journal of Information Science, and *Library Trends.* Citations were verified directly by comparison of the original publication or from a photocopy of the original publication.

The citations were classified by category of Internet resource. These citations were recorded and then the cited electronic resource was accessed on the Internet. The data collection form identified each individual electronic citation and the results were coded. The coding fields used were 1) citation category, 2) year of journal publication, 3) whether the item was successfully retrieved, 4) major errors, and 5) minor errors. If the item was found at the reference location given, it was recorded as accurate and stable. If the desired item was not found at the reference location given, the citation was noted and then analyzed. Next, distinctions were made between major and minor errors. Final the citations to inaccessible resources were studied to determine why an item was not retrieved.

The types of errors by category of electronic resource, by journal and by year are identified and recorded. The categories of electronic resources are based on a review of Internet-based resources as presented to ASLIB, (Fletcher and Greenhill 1995) and are coded as follows: FTP = FTP URL; Gopher= Gopher URL; HTTP = HTTP URL; and Listserv = listserv message. The errors are then classified as minor (see Table 1) or major (see Table 2) errors following Doms (1989) definitions.

Table 1: Minor error types

Error Type	Definition
Type A	Automatic URL address redirect
Type B	Recognizable error in address syntax
Type C	Address change with forwarding address provided; request resent
Type D	Wrong subdirectory; source located by conducting site search
Type E	Wrong subdirectory; source located by following link from page

Table 2: Major error types

Error Type	Definition
Type 1	Missing information (insufficient information to retrieve the source, e.g., Listserv name but no address)
Type 2	Host server not responding
Type 3	File not found (Error: 404)
Type 4	Unknown server or host message, no DNS registration
Type 5	Password required for access
Type 6	Access to archive restricted (e.g., to members only); do not have permission
Type 7	Document moved to another server or subdirectory; no link provided; page changed; no forwarding address
Type 8	Cannot access directory or server; no further information provided
Type 9	Bad request or malformed URL (the browser sent a query that the server could not understand); host server error

Results and discussion

In the ten library and information science journals examined for the years 1994-1998, 1157 citations to electronic resources were identified. Of those 1157 citations, only 637 items, or 55.1%, were retrieved. The rate of retrieval of

BEST COPY AVAILABLE

resources increases from 31.8% for 1994 to 66.5% for 1998. (See Table 3) All of the citations were verified by comparing the citation to the original article reference, or a photocopy of the original article. The electronic resources were classified by category of resource and analyzed for their individual retrieval rates. Items which used the HTTP URL were the most stable. They had the highest retrieval rate, 576 items out of 957, or 60.2%. References to Listserv postings had the lowest retrieval rate, 11 out of 83, or 13.3%. (See Table 4) It should be noted that the largest number of citations used the HTTP URL , 82.7%. As the World Wide Web continues to grow, this number may reflect the waning use of the older types of protocols, FTP, Gopher and Listservs, in favor of the newer HTTP protocol references. The lower retrieval rate among these older kinds of URLs may also be indicative of inconsistencies in the format of the citations used to reference them.

The four categories of references were also compared by year to determine if an individual category showed significant rate increases or decreases over time. (See Table 5) In general, FTP URLs declined from a retrieval rate of 77.8% in 1994 to 15.4% in 1998. In contrast, HTTP URLs increased in the same time period, from a retrieval rate of 0.0% in 1994 to 68.8% in 1998.

Table 3: Total citations retrieved by year

Year	Total number of citations	Total citations retrieved	Percentage of citations retrieved
1994	66	21	31.8%
1995	125	43	34.4%
1996	136	63	46.3%
1997	313	166	53.0%
1998	517	344	66.%
Totals	1157	637	*55.1%

*average rate of retrieved items for 1994-1998

Table 4: Total citations retrieved by reference category

Category	Total number of citations	Total citations retrieved	Percentage of citations retrieved
Listserv	83	11	13.3%
Gopher	41	17	41.5%
FTP	76	33	43.4%
HTTP	957	576	60.2%
Totals	1157	637	*55.1%

*average rate of retrieved citations

BEST COPY AVAILABLE

Table 5: Retrieval rate by category over time

Resource type: *FTP*

Year	Total number of citations	Total citations retrieved	Percentage of citations retrieved
1994	18	14	77.8%
1995	7	3	42.9%
1996	9	3	33.3%
1997	29	11	37.9%
1998	13	2	15.4%
Totals	76	33	*43.4%

Resource type: *Gopher*

Year	Total number of citations	Total citations retrieved	Percentage of citations retrieved
1994	1	0	0.0%
1995	24	13	54.2%
1996	8	0	0.0%
1997	3	2	66.7%
1998	5	2	40.0%
Totals	41	17	*41.5%

Resource type: *HTTP*

Year	Total number of citations	Total citations retrieved	Percentage of citations retrieved
1994	9	0	0.0%
1995	71	23	32.4%
1996	109	60	55.0%
1997	274	153	55.8%
1998	494	340	68.8%
Totals	957	576	*60.2%

Resource type: *Listserv*

Year	Total number of citations	Total citations retrieved	Percentage of citations retrieved
1994	38	7	18.4%
1995	23	4	17.4%
1996	10	0	0.0%
1997	7	0	0.0%
1998	5	0	0.0
Totals	83	11	*13.3%

*average rate of retrieved items for 1994-1998

BEST COPY AVAILABLE

In addition to examining the retrieval rates of the different categories of electronic resources being cited in the print literature, the procedure for this study also recorded and evaluated the number and types of errors in the citations. The total number of citations which included an error (either major or minor) was 625 (out of 1157) or 54.0%. (See Tables 6 and 7) For comparison, Pope's 1992 study revealed that 30% of the bibliographical references to print sources that were examined in ten library science journals contained errors. (Pope 1992)

The percentage of total errors for each resource category was calculated. The lowest rate of error was recorded for HTTP sources, 51.3%. The highest rate of error was for Listserv references at 71.1%.

Table 6: Total errors by resource category

	Total citations	Total minor errors	Total major errors	Total errors (%)
FTP	76	10	41	51 (67.1)
Gopher	41	0	24	24 (58.5)
HTTP	957	121	370	491 (51.3)
Listserv	83	1	58	59 (71.1)
Totals (%)	1157	132 (11.4)	493 (42.6)	625 (54.0)

BEST COPY AVAILABLE

Table 7: Total errors over time

	Total citations	Total minor errors	Total major errors	Total errors (%)
1994	66	4	36	40 (60.6)
1995	125	8	81	89 (71.2)
1996	136	19	65	84 (61.8)
1997	313	35	141	176 (56.2)
1998	517	66	170	236 (45.6)
Totals (%)	1157	132 (11.4)	493 (42.6)	625 (54.0)

The total number of citations found to contain error(s) were examined over time to observe if the percentage of such citations increased or decreased. The percentage of total errors for all citations was observed to increase over time as the citations aged. That is, the percentage of error rate for 1998 was 45.6%; 56.2% for 1997, 61.8% for 1996, and 71.2% for 1995. It was observed however, that the percentage of error for 1994 was 60.6%, 10.6% less than in 1995. No reason for this anomaly could be identified using the existing data. (See Table 7)

The citations were sorted by year and by category of URL to observe how many major and minor errors were present for each year studied. The citations were examined for major errors within the various URL categories. The largest percentage of major errors were identified for the Listserv references, 69.9%. The percentage of major errors for Gopher sites was 58.5%; for FTP sites, 53.9%; and for HTTP references, 38.7%. The percentage of errors for HTTP references is interesting because HTTP references account for the largest number of citations that were examined, 957 of 1157 total references, or 82.7%.

Of the 1157 items cited, it was observed that 132, or 11.4%, contained minor errors and 493 items, or 42.6%, contained major errors. The percentage of major errors, 42.6% (those errors that impede retrieval) is higher than the percentage of 30% Pope found in his study for all errors in print sources. (Pope 1992)

Five minor errors were found in the citations. These errors did not prevent retrieval of the electronic resource cited. HTTP URLs contained 121 of the 132 recorded minor errors, or 91.7%. (See Table 8) The prevalence of Type A errors (automatic URL redirect), which represents 53.8% of minor errors would seem to indicate that there is an increase use of the redirect to assist document retrieval. References with Type A error might not be retrieved without the use of the automatic redirect.

Table 8: Minor errors by citation category and type of error
n=132

Error Type	FTP	Gopher	HTTP	Listserv	Total by error type (%)
Type A	0	0	71	0	71 (53.8)
Type B	6	0	20	0	26 (19.7)
Type C	0	0	19	1	20 (15.2)
Type D	4	0	4	0	8 (6.1)
Type E	0	0	7	0	7 (5.3)
Total errors by category (%)	10 (7.6)	0 (0.0)	121 (91.7)	1 (0.7)	132 (100)

Finally, minor errors were found to increase steadily over time, from 3.0%

BEST COPY AVAILABLE

in 1994, to 50.0% in 1998. (See Table 9) The greatest increase occurred among those citations with a Type A error, 0 recorded for 1994 and 38 recorded for 1998. The total number of minor errors is approximately doubled each year. It would appear that the number of minor errors will continue to increase for 1999. It is uncertain however, if these types of errors will continue to be minor ones. That is, will changes in access methods of the different categories of URLs result in an increase in items not retrieved? If current Internet resources do not continue to leave "forwarding addresses", they may not be retrievable in the future.

Table 9: Minor errors by type over time

Error Type	1994	1995	1996	1997	1998
Type A	0	4	9	20	38
Type B	0	1	3	5	17
Type C	1	2	7	2	8
Type D	3	1	0	4	0
Type E	0	0	0	4	3
Total errors by year (%)	4 (3.0)	8 (6.1)	19 (14.4)	35 (26.5)	66 (50.0)

Nine types of major errors i.e., those errors that impeded retrieval of the requested electronic resource, were also identified. The most common major error was a Type 3 error (file not found, i.e., an Error: 404 message). Of the 493 total citations with major errors, 227, or 46.0% were of Type 3. (See Table 10) This number may indicate the growing number of citations found that utilized the

HTTP URL. (The 404 message is one of the automatic error messages generated when the HTTP address becomes non-functioning.)

In addition, 77 of the 493 total citations, or 15.6% were of Type 2 error, "server not responding." Repeated attempts were made on three separate days to access these URLs. Those for which the server continued to be non-responsive were then assigned to Type 2. Major errors related to host server problems (major error types 2, 4, 7, 8, and 9) accounted for 200 of the 493 total major errors, or 40.6%. This represents a significant portion of non-retrieved items for which standardization of citations will not affect. However, use of OCLC's PURL resolver service might be utilized to correct for server error by maintaining the associations that are lost when documents are moved to new locations.

Table 10: Major errors by citation category and type of error
n=493

Error type	FTP	Gopher	HTTP	Listserv	Total by error type (%)
Type 1	1	0	13	36	50 (10.1)
Type 2	13	11	46	7	77 (15.6)
Type 3	11	0	213	3	227 (46.0)
Type 4	5	5	53	11	74 (15.0)
Type 5	0	0	4	0	4 (0.8)
Type 6	2	0	9	1	12 (2.4)
Type 7	0	0	22	0	22 (4.7)
Type 8	9	8	2	0	19 (3.9)
Type 9	0	0	8	0	8 (1.6)
Total errors by citation category	41 (8.3)	24 (4.9)	370 (75.1)	58 (11.8)	493 (100)

BEST COPY AVAILABLE

The citations were sorted by year to observe how many major errors were present for each year studied. It was found that the percentage of major errors did increase over time from 7.3% in 1994 to 34.5% in 1998. In other words, 34.5% of all electronic resources cited for 1998 failed to be retrieved due to a major error. (See Table 11) This was not a steady increase however, due to the lower error rate observed for 1996. Again, no reason for this anomaly could be identified using the existing data.

While the rate of major error types 3 and 4 show an increase from 1994 to 1998 (with unaccounted anomalies for the year 1996), it should be noted that the rate of major error type 1 (missing information) actually decreased from 18 of 36, (or 50%) of the major errors identified for 1994 to 4 of 170, (or 2.35%) of the major errors identified for 1998.

Table 11: Major errors by error type over time
n=493

Error type	1994	1995	1996	1997	1998	Totals (%)
Type 1	18	12	8	8	4	50 (10.1)
Type 2	9	13	12	21	22	77 (15.6)
Type 3	1	37	23	72	94	227 (46.0)
Type 4	7	10	9	20	28	74 (15.0)
Type 5	0	0	3	0	1	4 (0.8)
Type 6	0	1	0	2	9	12 (2.4)
Type 7	0	3	4	6	9	22 (4.7)
Type 8	1	5	4	7	2	19 (3.9)
Type 9	0	0	2	5	1	8 (1.6)
Total major errors by year	36 (7.3)	81 (16.4)	65 (13.2)	141 (28.6)	170 (34.5)	493 (100)

BEST COPY AVAILABLE

Finally, the major errors were re-examined and two classes of errors were formed. The first class comprise errors over which the author has some level of control, i.e., the Type 1 errors. The second class comprise errors over which the technology (or the individuals who maintain the technology) are at fault, i.e., Type 2, Type 4, Type 7, and Type 9 errors. These four types all involve problems with the host server. When these two classes were compared over the five year period, it was observed that the class 1 errors are declining (18 in 1994 to 4 in 1998) while the class 2 errors are increasing (16 in 1994 to 60 in 1998). (See Table 12)

Table 12: Comparison of two classes of major errors over time

	1994	1995	1996	1997	1998	Totals
Class 1	18	12	8	8	4	50
Class 2	16	26	27	52	60	181

Conclusions

The results of this study would seem to suggest that numerous problems exist in citations to electronic resources. Unlike their print counterparts, these references are subject to new and different kinds of access problems and types of errors. The high number of sources that were not found would suggest that archiving and cataloging of these sources will become a growing necessity if the "scientific bricklaying" (Mitra 1970) of citation referencing is to remain on solid

BEST COPY AVAILABLE

ground.) It was observed that only 66% of electronic resources cited in 1998 were still available. This would seem to be a clear indication of how volatile these resources can be.

It would also appear that the types of errors that are appearing have changed over the years. Fewer errors related to inconsistencies in the format of citations were observed while errors related to computer server problems increased dramatically. This might suggest the growing recognition among authors to use a standardized format for the citation of electronic resources. The errors related to computer server problem suggests that additional technological standards are needed. One solution that has been proposed is OCLC's Persistent URL (PURL) resolution service. This service limits the errors that might arise as documents are moved from one server location to another by maintaining an intermediate association between a resource and its Internet address. Additional solutions may involve the archiving of these documents by research institutions, a function that is widely practiced for print resources.

More studies are needed to examine the growing number of problems in accessing and retrieving documents from the Internet. Increased recognition of the importance of standardization of references in the professional literature, the preservation of items being referenced, and improved accessibility to the scientific threads that become available on the World Wide Web are some of the areas which need to be further explored.

Bibliography

Baird, Laura M. and Charles Oppenheim. 1994. Do citations matter? *Journal of Information Science,* 20 (1), p. 2-15.

Berners-Lee, Tim. 1994. Universal Resource Identifiers in WWW: A Unifying Syntax for the Expression of Names and Addresses of Objects on the Network as used in the World-Wide Web. [Online] CERN. Available: http://www.w3.org/Addressing/rfc1630.txt [1998, September 7]

Casey, Carol. 1998. The cyberarchive: A look at the storage and preservation of web sites. *College & Research Libraries,* (July), p. 304-310.

Daigle, Leslie, Ron Daniel Jr. and Cecilia Preston. 1998. Uniform resource identifiers and online serials. *The Serials Librarian* Vol. 33, No. 3/4, p. 325-341.

Doms, C. A. 1989. A survey of reference accuracy in five national dental journals. *Journal of Dental Research* 68 (March), p. 442-444.

Fletcher, Gordon and Anita Greenhill. 1995. Academic referencing of internet-based resources. *ASLIB Proceedings*, vol. 47, no. 11/12, November/December, p. 245-252.

Ford, Charlotte E. and Stephen P. Harter. 1998. The downside of scholarly electronic publishing: Problems in accessing electronic journals through online directories and catalogs. *College & Research Libraries*, (July), p. 335-346.

Garfield, Eugene. 1979. Citation Indexing: Its Theory and Application in Science, Technology and Humanities. New York: Wiley.

Greenhill, Anita. 1998. Electronic References & Scholarly Citations of Internet Sources. [Online] Information Quality WWW Virtual Library. Available: http://www.gu.edu.au/gint/WWWVL/OnlineRefs.html [1998, August 16]

Guffey, Mary Ellen. 1997. APA Style Electronic Formats. *Business Communication Quarterly*, (March), p. 59-76.

Harter, S.P. and H.J. Kim. 1996a. The impact of electronic journals on scholarly communications: A citation analysis. [Online] ASIS. Available: http://info.lib.uh.edu/pr/v7/n5/hart7n5.html [1998, September 2]

Harter, Stephen P. and Hak Joon Kim. 1996b. Accessing electronic journals and other e-publications: An empirical study. *College & Research Libraries*, (September), p. 440-456.

Hoemann, George. 1995 Electronic Style--Internet Addresses: URIs, URLs, URNs. [Online] University of Tennessee at Knoxville. Available: http://funnelweb.utcc.utk.edu/~hoemann/address.html [1998, August 19]

ISO. 1998. *ISO 690-2, Bibliographic references to electronic documents. (excerpts)* [Online] Available: http://www.nlc-bnc.ca/iso/tc46sc9/standard/690-2e.htm [1998, August 19]

Li, Xia and Nancy Crane. 1996. Electronic Styles: A Handbook for Citing Electronic Information. 2nd. ed. Medford, NJ: Information Today.

Library of Congress. 1997. The Relation between URNs, Handles, and PURLs. [Online] Library of Congress. Available: http://learning.loc.gov/ammem/award/docs/PURL-handle.html [1998, November 2]

Mitra, A.C. 1970. The bibliographical reference: A review of its role. *Annals of Library Science and Documentation*, 17 (3-4), p. 117-123.

Moed, H.F. and M. Vriens. 1989. Possible inaccuracies occurring in citation analysis. *Journal of Information Science* 15: 95-107.

Pandit, Idrisa. 1993. Citation errors in library literature: A study of five library science journals. *Library & Information Science Research* 15: 185-198.

Parrott, Jim. 1998. URL-Stability Index for the Scholarly Societies Project. [Online] Scholarly Societies Project Available: http://www.lib.uwaterloo.ca/society/URL_stability_index.html [1998, August 17]

Pope, Nancy N. 1992. Accuracy of references in ten library science journals. *RQ* 32, no. 2 (Winter), p. 240-43.

Price, Derek J. De Solla. 1963. Little Science, Big Science. New York: Columbia University Press

Shafer, Keith, Stuart Weibel, Erik Jul, and Jon Fausey. 1996. Introduction to Persistent Uniform Resource Locators. [Online] OCLC Online Computer Library Center, Inc. Dublin, OH. Available: http://purl.oclc.org/OCLC/PURL/INET96 [1998, August 15]

Sollins, Karen and Larry Masinter. 1994. Functional Requirements for Uniform Resource Names. [Online] Available: http://info.internet.isi.edu:80/in-notes/rfc/files/rfc1737.txt [1998, August 15]

SSCI: Social sciences citation index journal citation reports. Philadelphia: Institute for Scientific Information, 1996.

Sweetland, James H. (1989). Errors in bibliographic citations: A continuing problem. *The Library Quarterly* 59, no.4 (October), p. 291-304.

Weibel, Stuart, Erik Jul, and Keith Shafer. 1996. PURLs: Persistent Uniform Resource Locators. [Online] OCLC Online Computer Library Center, Inc. Dublin, OH. Available: http://purl.oclc.org/OCLC/PURL/SUMMARY [1998, November 2]

Whitley, Peggy. 1997. Citing Online Sources, Internet and Westlaw, MLA Style. [Online] Available: http://www.nhmccd.cc.tx.us/contracts/lrc/kc/mla-internet.html [1998, August 16]

Zhang, Yin. 1998. The impact of Internet-based electronic resources on formal scholarly communication in the area of library and information science: A citation analysis. *Journal of Information Science*, 24 (4), p. 241-254.